

The 5th Machine Translation Marathon in the Americas

College Park, MD
May 28 – June 1 2019



Machine Translation Marathon



2007 Edinburgh, Scotland

2008 Wandlitz, Germany

2009 Prague, Czech Republic

2010 Dublin, Ireland

2010 Le Mans, France

2011 Trento, Italy

2012 Edinburgh, Scotland

2013 Prague, Czech Republic

2014 Trento, Italy

2015 Prague, Czech Republic

2016 Prague, Czech Republic

2017 Lisbon, Portugal

2018 Prague, Czech Republic

Machine Translation Marathon in the Americas



2015 Urbana-Champaign, Illinois



2016 Notre Dame, Indiana



2017 Dayton, Ohio



2018 Pittsburgh, PA



2019 College Park, MD

<http://statmt.org/mtma19>

Today's schedule

09:00-09:30	Welcome and Introduction to Machine Translation - Marine Carpuat
09:30-10:30	Lecture: Deep learning and computation graphs - Antonios Anastasopoulos
10:30-11:00	Coffee break
11:00-12:00	Project Proposals and Team Allocation
12:00-01:30	Lunch
01:30-02:30	Lecture: Evaluation and Quality Metrics - Christian Federmann
02:30-	Projects
03:00	Coffee

Machine Translation

A Very Brief Introduction

Marine Carpuat

With slides by Lane Schwartz and Philipp Koehn

1947

When I look at an article in Russian, I say to myself: This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.



Warren Weaver

1952 the first Conference on Machine Translation at MIT

Some of the Participants

- Yehoshua Bar-Hillel (MIT)
- Andrew Booth (Birkbeck College, London)
- Leon Dostert (Georgetown University)
- Erwin Reifler (University of Washington)
- Victor Oswald (UCLA)

Topics Discussed

- OCR
- Document summarization
- Limitations of computing hardware
- Pre-editing, post-editing, and constrained language
- Problem of translating idioms
- Domain adaptation
- Out-of-vocabulary words and Zipfian distributions
- Assumed importance of morphology

1950s-1960s

- 1954 Georgetown-IBM experiment
 - 250 words, 6 grammar rules
- 1966 ALPAC report
 - Skeptical in research progress
 - Led to decreased US government funding for MT

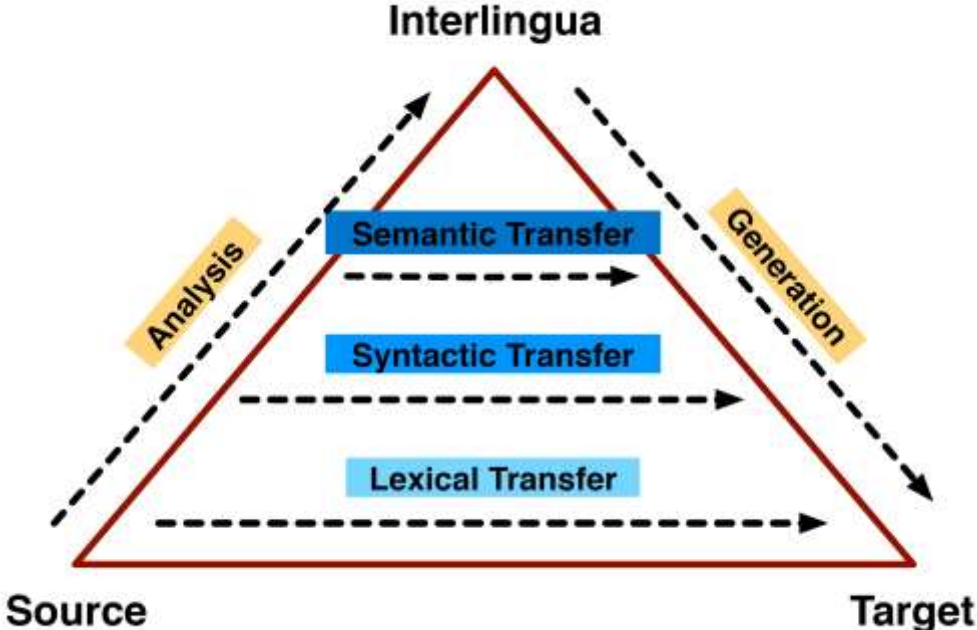


Rule based systems

- Approach
 - Build dictionaries
 - Write transformation rules
 - Refine, refine, refine
- Meteo system for weather forecasts (1976)

```
"have" :=  
  
if  
    subject(animate)  
    and object(owned-by-subject)  
then  
    translate to "kade... aahe"  
if  
    subject(animate)  
    and object(kinship-with-subject)  
then  
    translate to "laa... aahe"  
if  
    subject(inanimate)  
then  
    translate to "madhye...  
aahe"
```

The Vauquois Triangle



1988

A STATISTICAL APPROACH TO MACHINE TRANSLATION

Peter F. Brown, John Cocke, Stephen A. Della Pietra, Vincent J. Della Pietra, Fredrick Jelinek,
John D. Lafferty, Robert L. Mercer, and Paul S. Roossin

IBM

Thomas J. Watson Research Center
Yorktown Heights, NY

In this paper, we present a statistical approach to machine translation. We describe the application of our approach to translation from French to English and give preliminary results.

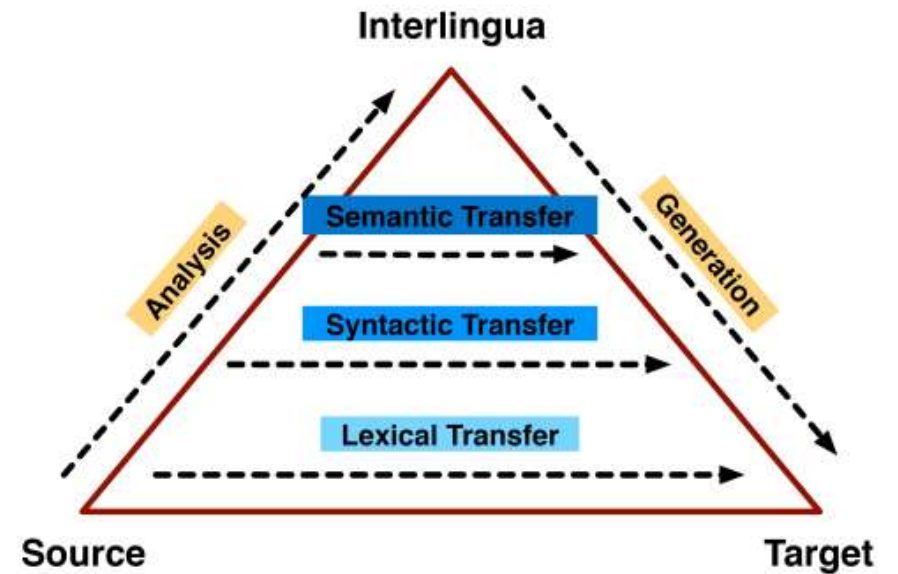
The COLING Paper Review

The validity of statistical (information theoretic) approach to MT has indeed been recognized, as the authors mention, by Weaver as early as 1949. And was universally recognized as mistaken by 1950. (cf. Hutchins, MT: Past, Present, Future, Ellis Horwood, 1986, pp. 30ff. and references therein) The crude force of computers is not science. The paper is simply beyond the scope of COLING.

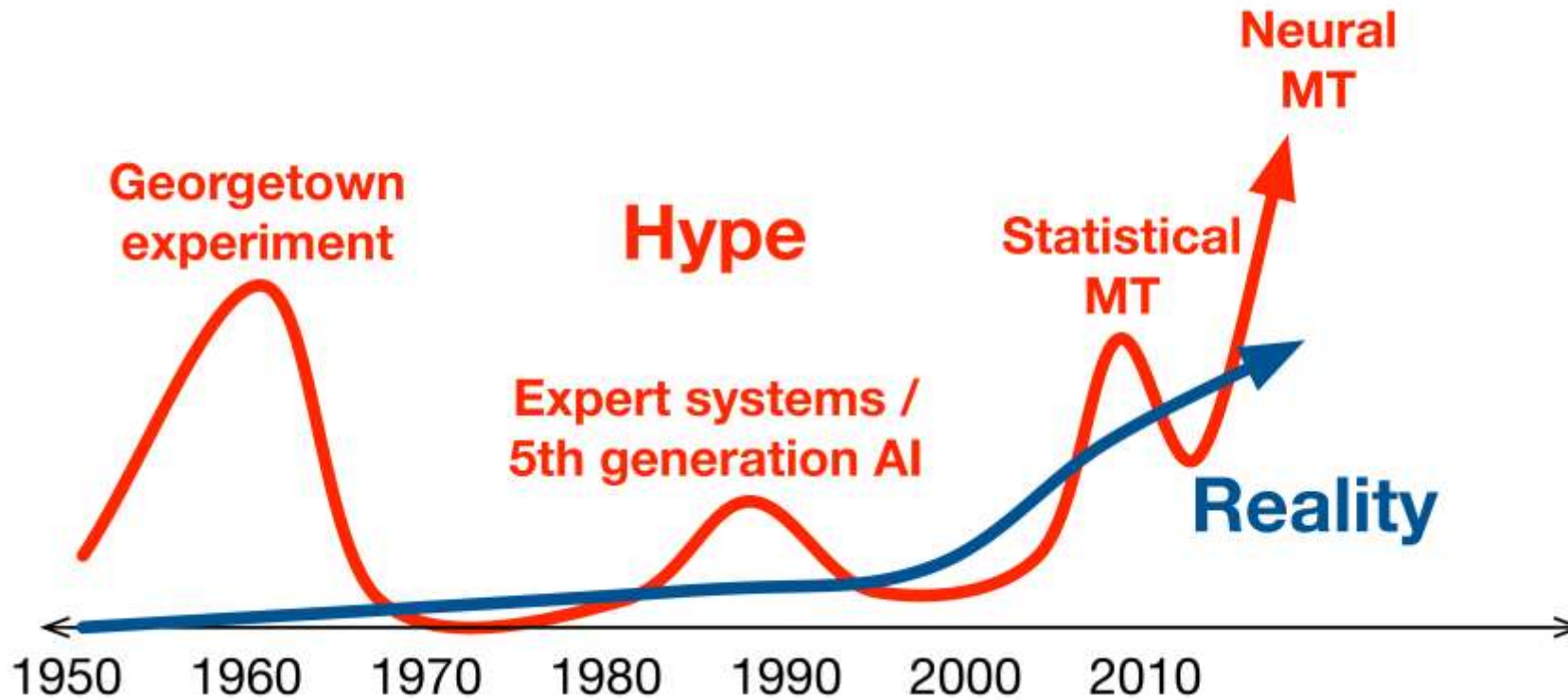
More about the IBM story: [20 years of bitext workshop](#)

Statistical Machine Translation

- 1990s: increased research
- Mid 2000s: phrase-based MT
 - Moses, Google Translate
- Around 2010: commercial viability
- Since mid 2010s: neural network models



MT History: Hype vs. Reality



How Good is Machine Translation Today?

March 14 2018:

Achieving Human Parity on Automatic Chinese to English News Translation

Hany Hassan*, Anthony Aue, Chang Chen, Vishal Chowdhary, Jonathan Clark,
Christian Federmann, Xuedong Huang, Marcin Junczys-Dowmunt, William Lewis,
Mu Li, Shujie Liu, Tie-Yan Liu, Renqian Luo, Arul Menezes, Tao Qin,
Frank Seide, Xu Tan, Fei Tian, Lijun Wu, Shuangzhi Wu, Yingce Xia,
Dongdong Zhang, Zhirui Zhang, and Ming Zhou

Microsoft AI & Research

How Good is Machine Translation Today?

Top scoring system output (BLEU) at WMT19 German-English

http://matrix.statmt.org/matrix/output/1902?run_id=6910&view_start=20

Source

Laut Besitzer der brennenden Gartenhütte hatte er zu Mittag einen Akku zum Laden angesteckt.

MT

According to the owner of the burning garden hut, he had infected a battery for charging at noon.

Reference

The owner of the burning garden hut stated that at noon he had connected a battery to the power supply for charging.

How Good is Machine Translation Today?

Top scoring system output (BLEU) at WMT19 German-English

http://matrix.statmt.org/matrix/output/1902?run_id=6910&view_start=20

Source

Das Dach musste praktisch zur Gänze entfernt und Brandschutt im unbedingt erforderlichen Ausmaß ausgeräumt werden.

MT

The roof had to be removed practically in its entirety and fire debris had to be removed to the extent absolutely necessary.

Reference

The roof had to be almost completely removed and the huge amount of debris needed clearing.

How Good is Machine Translation Today?

Top scoring system output (BLEU) at WMT19 Gujarati-English

http://matrix.statmt.org/matrix/output/1904?run_id=6756

Source

પતંજલી લોટના પેકેટમાં
નિકળ્યો દેડકો, સુપર
બાજારથી ખરીદ્યું હતું લોટ

મધ્યપ્રદેશના સીહોરમાં
પંતજલીના લોટ પેકેટમાં મરેલું
દેડકો મળવાની ખબર આવી
છે.

MT

Nick Delgado in Patagonia
Lodge's pack, bought from
Super Bazaar.

It has been reported to have
found dead frogs in
Pantjali's flour pack in the
Middle East.

Reference

Frog inside Patanjali 'Aata'
(flour) packet, was
purchased from the Super
Bazaar.

There is news that a dead
frog was found in a Patanjali
Aata packet in Sehore
district of Madhya Pradesh.

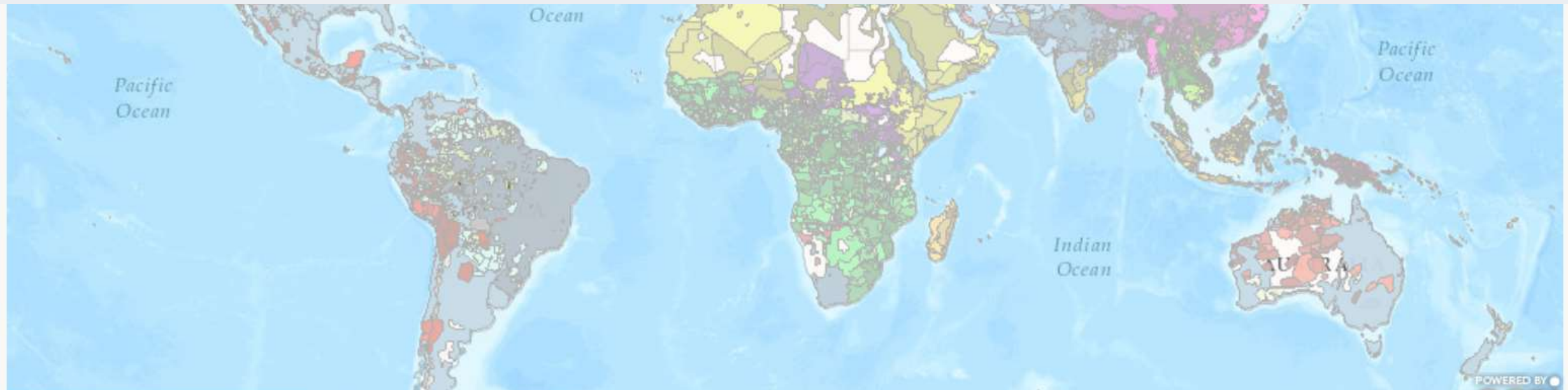
How Good is Machine Translation Today?



<https://www.haaretz.com/israel-news/palestinian-arrested-over-mistranslated-good-morning-facebook-post-1.5459427>

Search for a language, dialect name or major city...

6,800 living languages
600 with written tradition
100 spoken by 95% of population



Machine Translation

A Very Brief Introduction

Marine Carpuat

With slides by Lane Schwartz and Philipp Koehn